

AD _____

Award Number: DAMD17-02-1-0364

TITLE: Genome-Wide Nucleic Acid/Protein Interactions in Breast Cancer

PRINCIPAL INVESTIGATOR: Pamela A. Silver, Ph.D.

CONTRACTING ORGANIZATION: Dana-Farber Cancer Institute
Boston, MA 02115

REPORT DATE: May 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20030902 143

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY
(Leave blank)**2. REPORT DATE**
May 2003**3. REPORT TYPE AND DATES COVERED**
Annual (15 Apr 2002 - 14 Apr 2003)**4. TITLE AND SUBTITLE**

Genome-Wide Nucleic Acid/Protein Interactions in Breast Cancer

5. FUNDING NUMBERS

DAMD17-02-1-0364

6. AUTHOR(S)

Pamela A. Silver, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)Dana-Farber Cancer Institute
Boston, MA 02115

E-Mail: Pamela_silver@dfci.harvard.edu

**8. PERFORMING ORGANIZATION
REPORT NUMBER****9. SPONSORING / MONITORING
AGENCY NAME(S) AND ADDRESS(ES)**U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012**10. SPONSORING / MONITORING
AGENCY REPORT NUMBER****11. SUPPLEMENTARY NOTES**

Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE**13. ABSTRACT (Maximum 200 Words)**

Since many types of breast cancer remain untreatable, the research proposal aims to develop novel genomic technology to identify potential therapeutic targets and to aid in diagnosing various types of breast cancer at the molecular level. The overarching goal of the proposal is to develop a technology to screen nucleic-acid protein interactions on a genome scale with a focus on understanding complexes involved in breast cancer. In order to identify the regulatory networks of interactions between RNAs and proteins, we propose to develop a rapid genome-scale method to determine the specific RNA targets and RNA binding sites of proteins. The aims were to 1) discover RNA targets of specific RNA binding proteins and protein complexes including the breast cancer susceptibility gene BRCA1, and 2) define the RNA sequences recognized by proteins using novel nanotechnologies including development of optically encoded beads containing both a unique optical signature and a specific oligonucleotide. We have completed the proof of concept experiments for the bead-based assay, identified RNAs that may interact with a BRCA1 complex and are complementing the technology with genome-wide chromatin immunoprecipitation.

14. SUBJECT TERMS

RNA-protein interactions; DNA-protein interactions; genomic screening; regulation of gene expression; technology development

15. NUMBER OF PAGES

33

16. PRICE CODE**17. SECURITY CLASSIFICATION
OF REPORT**

Unclassified

**18. SECURITY CLASSIFICATION
OF THIS PAGE**

Unclassified

**19. SECURITY CLASSIFICATION
OF ABSTRACT**

Unclassified

20. LIMITATION OF ABSTRACT

Unlimited

Table of Contents

Cover.....	1
SF 298.....	2
Table of Contents.....	3
Introduction.....	4
Body.....	4
Key Research Accomplishments.....	8
Reportable Outcomes.....	8
Conclusions.....	8
References.....	9
Appendices.....	9

INTRODUCTION:

Since many types of breast cancer remain untreatable, the research proposal aims to develop novel genomic technology to identify potential therapeutic targets and to aid in diagnosing various types of breast cancer at the molecular level. The overarching goal of the proposal is to develop a technology to screen nucleic-acid protein interactions on a genome scale with a focus on understanding complexes involved in breast cancer. In order to identify the regulatory networks of interactions between RNAs and proteins, we proposed to develop a rapid genome-scale method to determine the specific RNA targets and RNA binding sites of proteins. The aims were to 1) discover RNA targets of specific RNA binding proteins and 2) define the RNA sequences recognized by proteins using novel nanotechnologies including development of optically encoded beads containing both a unique optical signature and a specific oligonucleotide. A labeled genomic RNA pool will be bound to a labeled protein and then hybridized to the oligonucleotide-bead library. Beads with bound protein are selected by flow cytometry and the oligo sequences decoded and assembled to identify the bound RNA. This technology is being complemented by genome-wide chromatin immunoprecipitation. Progress has been made on Aims 1-3 of the original grant as detailed below.

BODY:

Aim 1 was successfully completed as marked by the publication of a paper describing the assay (Brodsky and Silver, 2002). We are currently progressing with Aim 3. Although the goals remain the same, we are using new, more powerful, methods than originally proposed in Aim 2. Thus, we have laid the groundwork for genomic and small molecule screening using the microbead assay. The assay has generated some interest in the community as we have recently written an invited review discussing the microbead assay we have developed (Brodsky et al., 2003). The research accomplishments associated with each task outlined for the first year of funding in the approved Statement of Work are detailed below.

Technical Objective 1: Determine optimal conditions for bead-based genomic screening

TASK 1: Construct vectors for RNA transcription and protein expression. We are using Ambion's pDP T7 and T3 polymerase plasmids for RNA transcription. We are using Invitrogen's gateway system to clone and purify GST and his tagged proteins. Because a number of labs and Invitrogen are pursuing efforts to clone and verify human and mouse ORFs into this recombination cloning system, it will be flexible and save us significant cloning effort in the future.

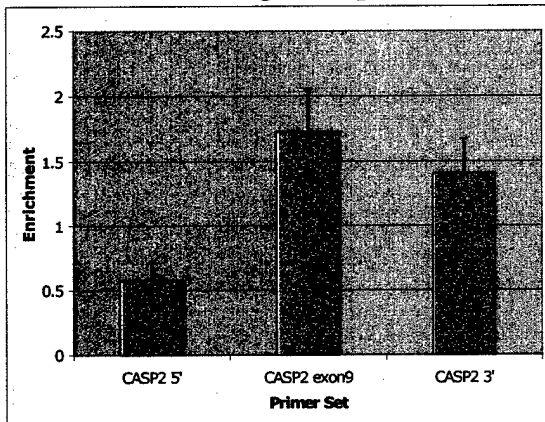
TASK 2: Determine optimal method for purifying fluorescently labeled protein. This is one of our primary focuses at this time. We are trying fluorescently labeled anti-GST and anti-his antibodies as well as in vitro labeling methods. The GFP fluorophore was convenient to develop the assay with the U1A system; however, other fluorophores such as fluorescein are approximately 30 times brighter and may lead to higher sensitivity and dynamic range of the assay. We are purifying other U1A constructs to compare to GFP and develop this technology.

TASK 3. Determine conditions for end-labeling genomic RNA. The "perfect" 3' end-labeling would uniformly label a wide-variety of RNA species. Each of the commonly used enzymes, poly(A) polymerase, ligase, etc. have certain drawbacks. Although poly(A) polymerase would be a strong and efficient labeling system for mRNAs. We are currently testing the micromax labeling system from NEN that post-transcriptionally body labels the RNA. We expect this to give a uniform and high sensitivity signal.

Technical Objective 2: Identification of target RNAs of clinically important proteins

TASK 1: Express and purify proteins such as AUF1 and HuR as fusion proteins in either *E. coli* or insect cells. We have constructed and sequence verified Gateway clones of 3 Hu proteins and are close to having purified protein suitable for the assay. This process will continue with the other Hu proteins once Objective 1, Task 2 is completed.

Task 2: Search for RNA targets by comparing complexes between breast cancer and normal cells. While the new microbead technology is being developed, current microarray technologies could be used to determine candidate binding targets. Recently, Chromatin Immunoprecipitation (ChIP) has emerged as a powerful method to identify where on a gene and - in combination with microarrays (ChIP-chip) - on which genes chromatin associated proteins are binding (Lei et al., 2001; Lei and Silver, 2003). Briefly, cells are cross-linked and chromatin is sheared to approximately 400 bp average size. The protein of interest is immunoprecipitated and the DNA is isolated for quantitative PCR analysis or microarrays. Because our lab, as well as others, has shown that many RNA binding proteins bind co-transcriptionally, we can take advantage of this approach. In addition representing a more *in vivo* situation, a potentially significant advantage of ChIP is that the cells are formaldehyde cross-linked allowing the capture of dynamic interactions.



Initially, we are focusing on multifunctional RNA binding proteins such as Polypyrimidine Binding Protein (PTB) and Aly. These proteins are known to regulate the alternative splicing and expression, respectively, of genes critical to cell growth and apoptosis such as caspases and fibroblast growth factors. We have tested a number of antibodies and have found that PTB and Aly can be co-immunoprecipitated with their target genes. genes more

Figure 1. Enrichment of PTB binding in chromatin is observed in the middle and end of the Caspase 2 gene compared to the 3' end. Data is normalized to a gene not highly expressed in HeLa cells.

than TFIID which appears 5' biased. These proteins are recruited to caspase2 and FGFR2 in the middle and 3' ends of the genes as compared to TFIID which appears 5' biased. Figure 1 shows some representative data of PTB recruitment. Importantly, we find that these proteins are recruited to the 3' ends such that the recruited DNA could be probed on the currently available 3' biased cDNA arrays. We are currently performing these array experiments. We are analyzing the proteins' recruitment in cell lines such as HeLa and comparing the results to cell lines relevant to breast cancer such as MCF7. In sum, we are using ChIP to analyze the recruitment of RNA binding proteins across particular genes to understand the mechanism while also determining the genome-wide recruitment with microarrays. The goal is to connect these RNA binding and splicing events to various cell signaling and cell response pathways to understand how cells become cancerous at the post-transcriptional level.

We have also extended our analyses to determine the nature of the predicted interaction between BRCA1, RNA and heterochromatin. Prompted by the association of BRCA1 with XIST RNA and the suggestion that chromatin structure is dependent on RNA, we determined to propose experiments to analyze BRCA1-bound RNAs. Some preliminary results that suggest BRCA1 associates with small RNAs are presented below.

i. BRCA1-associated RNA immunoprecipitation

To investigate whether BRCA1 interacts with RNAs, BRCA1 was immunoprecipitated from 293 cell extracts with both mono and polyclonal anti-BRCA1 antibodies. The immunoprecipitations were performed under RNase-free conditions and extracts were treated with RNase inhibitors.

Following DNase treatment, BRCA1-bound RNA was eluted using proteinase K. The resulting supernatant was extracted with phenol, phenol:chloroform and chloroform, respectively, followed by ethanol precipitation in the presence of 20 ug glycogen. The pellet was washed twice with 70% ethanol and resuspended in small volume of water.

ii. 3'-end labeling of immunoprecipitated RNA

Immunoprecipitated RNA was 3'-end labeled using T4 RNA ligase and [5'-³²P]-cytidine 3',5'-bis(phosphate) (pCp) in the presence of DMSO. The labeled products were purified and resolved on a 15% TBE/urea polyacrylamide gel and subject to autoradiography.

iii. BRCA1-associated RNA immunoprecipitated RNA

Immunoprecipitated RNA was resolved on a 15% TBE/urea polyacrylamide gel in the presence of trace radiolabeled 21-nt RNA marker. The radioactive band was excised, recovered from the gel and precipitated. A 3' RNA adapter, containing a 5' phosphate and 4 5' terminal ribonucleotides, followed by a stretch of 17 deoxynucleotides (with the most 3' terminal sugar being a dideoxynucleotide to prevent adapter-adapter ligation) was added to the RNA by an RNA ligation reaction. The ligation products were purified on a 12% polyacrylamide Urea/TBE gel and were then subject to a similar recovery as described above. A 5' RNA adapter was added and purified. The doubly ligated RNA species were reverse-transcribed and amplified by PCR. The resulting cDNAs were then concatamerized by virtue of an EcoRI restriction site with in the adapter sequence. Concatamers larger than 300 bp were selected and cloned into the pCR2.1-TA vector. Positive clones were isolated and the plasmids were sequenced. Deoxynucleotide oligomers were synthesized according to candidate RNA sequences and used as probes in Northern blot hybridizations to analyze the expression, processing and abundance of potential BRCA1-associated RNAs. Furthermore, various sequence analysis was performed on these RNAs, which included evolutionary conservation analysis, hairpin folding prediction and putative target prediction.

Figure 2 shows resolved 3'-end labeling reactions of BRCA1 immunoprecipitated RNA. Specific RNAs appear to associate specifically with BRCA1, as both polyclonal and monoclonal IP reactions, but not the preimmune control reaction, label 21-nucleotide RNAs. Five distinct RNAs were cloned from BRCA1 immunoprecipitations, ranging in size 23-28 nucleotides long. The expression of each RNA in total RNA extracted from 293 cells was examined by Northern blot analysis using 5'-³²P labeled oligonucleotide probes. 3 out of the 5 probes hybridize to RNAs on the Northern analysis. However, the sizes of the corresponding bands are not consistent with the cloned size or the size of the RNA observed from the 3'-labeling of immunoprecipitated RNAs. This RNA is present in different sizes, with the most abundant size in the 85-90 nucleotide size range. This result suggests that the RNA may exist as a larger precursor and their association with BRCA1 exists as the smaller processed form (21-nt). However, the observed 21-nt band observed from the labeling experiment may be a degradation product.

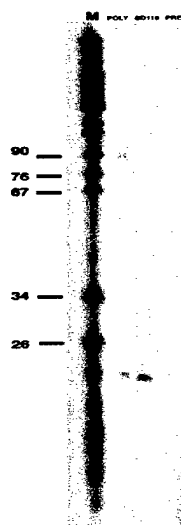


Figure 2: 3'-end labeling of BRCA1

Immunoprecipitated RNAs

Lane 1: Marker - pBR322 MspI digest

Lane 2: Polyclonal anti-BRCA1 antibody

Lane 3: SD118 monoclonal anti-BRCA1 antibody

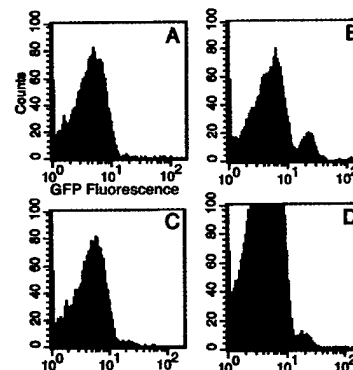
Lane 4: Preimmune control

Technical Objective 3: Determine binding sites by targeting RNA transcripts with bead libraries

Tasks 1- 2: Target known RNA-protein interactions such as U1A. We have completed the goals of these tasks.

The microbead assay accurately monitored the specificity and affinity of the U1 RNA-protein interaction. We also demonstrated that the complex could be identified even from total RNA. In addition to the published data, we have performed other experiments to establish the system for genomic screening such as the experiment in Figure 3. For screening RNA-protein interactions with oligonucleotide bead libraries, it will be important to detect binding when the beads containing the cognate oligonucleotide are rare in the bead population. Different dilutions of oligo 7 were mixed with the noncomplementary oligo 21 for U1A test RNA binding reactions with U1A-GFP. The results are summarized in Figure 3. These data also suggest the observed fluorescence shifts are sufficient to differentiate binding and non-binding bead populations. With multiplexing experiments, the beads will be distinguished by their code and not simply by the shifted population.

Figure 3. U1A-GFP binding to a bead populations in the presence of a large excess of other beads. Oligo 7 beads were diluted into oligo 21 beads, which are noncomplementary to the U1A test RNA. Binding reactions include 10 nM U1A RNA and 100 nM U1A-GP. 10,000 events are counted. **A)** Oligo 21 beads alone. **B)** 1:10 dilution of oligo 7 in oligo 21 beads. A distinct population of shifted beads is observed. **C)** 1:100 dilution of oligo 7 beads in oligo 21 beads. A shifted bead population is observed. **D)** The same 1:100 dilution reaction with 30,000 events counted. The shifted oligo 7 beads are now more clearly observed.



Task 3. Screen Model RNA-protein interactions with bead libraries in the presence of cell lysate. The first step towards screening is to develop some level of multiplexing. U1A binding experiments have been performed with coded beads from the Trau Lab. With the coded beads, we have performed multiplexing reactions with up to 4 oligos at once. These data demonstrated that the hybridization competition did not affect the RNA-protein binding results i.e. the oligos behaved as they did when tested one at a time.

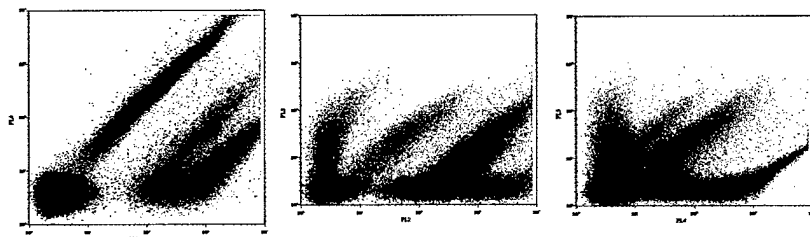


Figure 4. 8 coded beads from the Trau Lab are coded using 3 different fluorophores. From the 3 plots, each color representing a unique code can be differentiated from all the others.

Current work is focusing on improving the bead surface properties to minimize both the nucleic acids and protein nonspecific binding.

KEY RESEARCH ACCOMPLISHMENTS:

- Successful development of the bead-based assay using model proteins
- Determination of the co-transcriptional recruitment of certain RNA binding proteins in mammalian cells by chromatin immunoprecipitation allowing for genome-wide microarray analyses
- Identification of potential BRCA1-associated RNAs
- Development of multiplexing with coded beads

REPORTABLE OUTCOMES:

Brodsky, A. and Silver, PA. A microbead-based system for identifying and characterizing RNA-protein interactions by flow cytometry. 2002, Mol Cell Proteomics, 12:922-9

Brodsky, AS, Johnston, APR, Trau, M. and Silver, PA. Analysis of RNA-protein interactions by flow cytometry. Curr Op in Molecular Therapeutics. 2003, In press.

CONCLUSIONS:

Post-transcriptional regulation is crucial for proper growth regulation as evidenced by alternative mRNA splicing and 3' end formation, regulated export of RNAs out of the nucleus, repression at the level of translation and RNA degradation (Lei and Silver, 2002). It is clear that all of these processes impact on disease yet remain poorly understood. Our results begin to allow for a broader look at how these processes are altered in breast cancer. In particular, we have developed powerful new technologies to assess the behavior of key RNA binding proteins in both normal and cancer cells. We have made progress applying these assays and have preliminary data regarding the association of key RNA binding proteins with certain genes as well as the novel finding that BRCA1 may interact with certain RNAs. Taken together, the ability to assess on a genome-wide scale how proteins interact with

RNAs will address fundamental questions in biology including the regulation of gene splicing as well as potentially provide new therapeutic strategies.

REFERENCES:

Brodsky, A. and Silver, PA. A microbead-based system for identifying and characterizing RNA-protein interactions by flow cytometry. 2002, Mol Cell Proteomics, 12:922-9

Brodsky, AS, Johnston, APR, Trau, M. and Silver, PA. Analysis of RNA-protein interactions by flow cytometry. Curr Op in Molecular Therapeutics. 2003, In press.

Lei, EP, Krebber, H. and Silver, PA Messenger RNAs are recruited for nuclear export during transcription. 2001. Genes & Dev. 15, 1771-82.

Lei, E. and Silver, PA. Protein and RNA export out of the nucleus. Developmental Cell, 2002, 2:1-20.

Lei, E. and Silver, PA. Intron status and 3' end formation control cotranscriptional export of mRNA. Genes & Dev., 2002. 16:2761-2766.

APPENDICES:

Original copies of 2 relevant journal articles:

Brodsky A and Silver PA. A microbead-based system for identifying and characterizing RNA-protein interactions by flow cytometry. Molecular & Cellular Proteomics 2002; 1.12: 922-929.

Brodsky AS, Angus PR, Johnston MT, and Silver PA. Analysis of RNA-protein interactions by flow cytometry. Curr Op in Molecular Therapeutics 2003. In press.

A Microbead-based System for Identifying and Characterizing RNA-Protein Interactions by Flow Cytometry*

Alexander S. Brodsky†§ and Pamela A. Silver

We present a high throughput, versatile approach to identify RNA-protein interactions and to determine nucleotides important for specific protein binding. In this approach, oligonucleotides are coupled to microbeads and hybridized to RNA-protein complexes. The presence or absence of RNA and/or protein fluorescence indicates the formation of an oligo-RNA-protein complex on each bead. The observed fluorescence is specific for both the hybridization and the RNA-protein interaction. We find that the method can discriminate noncomplementary and mismatch sequences. The observed fluorescence reflects the affinity and specificity of the RNA-protein interaction. In addition, the fluorescence patterns footprint the protein recognition site to determine nucleotides important for protein binding. The system was developed with the human protein U1A binding to RNAs derived from U1 snRNA but can also detect RNA-protein interactions in total RNA backgrounds. We propose that this strategy, in combination with emerging coded bead systems, can identify RNAs and RNA sequences important for interacting with RNA-binding proteins on genomic scales. *Molecular & Cellular Proteomics* 1:922-929, 2002.

RNA-protein interactions are a central component of post-transcriptional regulation at multiple levels including RNA processing, transport, and translation. The sequenced human genome reveals hundreds of potential RNA-binding proteins (1). A critical step toward understanding the function of RNA-binding proteins is to identify and determine how they interact with their target RNAs.

Several strategies have been developed to identify RNA-protein interactions. Genetic approaches include three-hybrid screens (2), phage display (3), and TRAP (translational repression assay procedure) (4) to identify proteins that bind a specific RNA. However, these strategies are generally not applicable to larger RNAs and not suitable for determining binding constants. SELEX (systematic evolution of ligands by

exponential enrichment) can identify high affinity RNA sequences that may or may not reflect the biologically relevant recognition site (5).

Recently immunoprecipitation has been combined with microarray analysis to identify candidate RNAs bound to proteins (6, 7). This approach is very promising for inspection of RNA-protein interactions on a genome-wide scale. However, it relies on the ability to preserve stable interactions during immunoprecipitation; many potentially weak interactions may be lost. In addition, extensive motif searching together with additional experimentation may be required to validate the biological significance of any interactors (8).

Recent advances in bead coding technologies to create high complexity platforms are leading to the development of new approaches for high throughput screening studies that could be amenable to the study of RNA-protein interactions (9-12). In principle, nucleic acid hybridization on microbeads offers a number of advantages over DNA chips including shorter hybridization times and better control of binding conditions (13). Therefore, we have developed a new equilibrium binding method on microbeads for elucidating the recognition site of an RNA-binding protein on its cognate RNA. The approach uses oligonucleotide-coupled microbeads and fluorescently labeled RNAs and proteins to monitor RNA-protein binding by flow cytometry. To develop the system for screening RNA-protein interactions, we demonstrate how this approach can be used to identify and characterize the interaction between the spliceosomal protein U1A and a hairpin derived from U1 snRNA as well as detect specific RNAs from total RNA populations.

EXPERIMENTAL PROCEDURES

Plasmids—The U1A test transcript was constructed by annealing overlapping oligos and ligating the annealed product into pDP19 (Ambion) to create plasmid pPS2702. The oligo sequences are: AAT-TCTTTATCTTCAAAGTTGTCTGTCCAAGATTGGACTTGTCCGGAG-TGCAATGGACG, AAGGACAAGCGTGTCTTCATCAGAGTTGACTTC-ACTCGAG, GACAAGTCCAAATCTTGGACAGACAACCTTTGAAGATA-AAG, and GATCCTCGAGTGAAGTCAACTCTGATGAAGACACGCTT-GTCCTTCGTCATTGCACTCCG.

U1A-green fluorescent protein (GFP)¹ was PCR-amplified from pPS2035 and ligated into prSETB (Invitrogen) to create pPS2699. The 96A→G U1A point mutant was constructed by using Stratagene's QuikChange system to create pPS2703. 77C→G was constructed by

From the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School and The Dana-Farber Cancer Institute, Boston, Massachusetts 02115

Received, October 1, 2002, and in revised form, November 27, 2002

Published, MCP Papers in Press, December 1, 2002, DOI 10.1074/mcp.T200010-MCP200

¹ The abbreviation used is: GFP, green fluorescent protein.

ligating annealed oligonucleotides into pDP19 as described above. All constructs were verified by sequencing.

Transcription and RNA Preparations—PPS2702 and pPS2703 were linearized with *EcoRI* and subsequently transcribed with Ambion's T3 polymerase Maxiscript kit. Labeling with ^{32}P verified a product of the expected size, and subsequent transcription reactions were purified by G-50 spin columns or phenol extractions followed by multiple ethanol precipitations. Texas Red-5-UTP (Molecular Probes) was incorporated during transcription and purified with G-50 spin columns. Total RNA from HeLa cells was prepared by the TRIzol method with high salt precipitations to reduce background GFP fluorescence. Yeast RNA was isolated by a hot phenol method. RNA concentrations were determined by UV spectrometry.

U1A-GFP Purification—Cells were grown to 0.5 optical density before induction with 1 mM isopropyl-1-thio- β -D-galactopyranoside for 4 h. Cells were resuspended in 20 mM HEPES, 10 mM KCl, 0.1% IGEPAL and lysed with lysozyme followed by sonication. After centrifugation, lysate was applied to nickel columns, washed extensively, and eluted with imidazole. Green fractions were pooled and dialyzed into 10 mM HEPES, pH 7.6, 10 mM KCl, 0.1% IGEPAL. To remove the histidine tag, 1.25 units/ μl enterokinase were added and incubated for >48 h at 25 °C. Enterokinase was removed with EKaway resin (Invitrogen). U1A-GFP was dialyzed into storage buffer (20 mM HEPES, 20 mM KCl, 0.1% IGEPAL, 10% glycerol). Concentrations were determined by comparing U1A-GFP to bovine serum albumin on Coomassie gels and by the Bio-Rad protein assay. Protein stored at -80 °C bound RNA similarly to fresh preparations (data not shown).

Bead Preparation—Before coupling, Dynal 2.8- μm magnetic streptavidin beads (M-280) were vortexed and/or sonicated to reduce aggregation. Similar to Dynal's recommended protocols, oligonucleotides were attached to beads with 1 nmol of oligonucleotide/ 9×10^6 beads/30 μl . Incubations longer than 5 h were required to reach maximum oligonucleotide density (data not shown). Similar procedures were used for the SpheroTech 5.7- μm magnetic streptavidin beads. Oligonucleotides were synthesized with a 12-carbon spacer and 5' biotin from two different sources: Dana-Farber Cancer Institute Molecular Biology Core Facilities and Integrated DNA Technologies, Inc. Oligonucleotides from each source behaved similarly.

Bead Binding Assays—Binding was performed in 20 mM HEPES, pH 7.5, 300 mM KCl, 0.1% IGEPAL, 10 ng/ μl tRNA, 0.04 units/ μl superase-IN (Ambion), and 20 ng/ μl bovine serum albumin unless otherwise indicated. RNA was heated to 95 °C for 1 min and cooled on ice before being mixed with U1A-GFP for at least 20 min at room temperature before addition of 1×10^5 oligonucleotide-coupled beads. Binding reactions were incubated at 35 °C for at least 14 h with constant rotation unless otherwise indicated. Shorter incubations (<6 h) gave lower fluorescence.

Flow Cytometry and Data Analysis—A BD Biosciences Vantage was used to sort beads with both GFP and Texas Red signals. A FACScan was used to monitor GFP alone. Typically 5,000–10,000 beads were counted, and the peak channel, which indicates the maximum height of the bead population, is used to estimate the fluorescence intensity. To determine the percentage of the population shifted to higher fluorescence, cut-offs were set relative to the background fluorescence. Beads with fluorescence above the cut-off are counted in the shifted population. For the binding curves, pro Fit (Quantum Soft) was used to fit the fluorescence intensities to a Langmuir isotherm.

RESULTS

The experimental design for analysis of protein-RNA interactions uses oligonucleotides coupled to microbeads to probe RNA-protein interactions and is outlined in Fig. 1A. To carry out the analysis, a protein-RNA complex is first formed

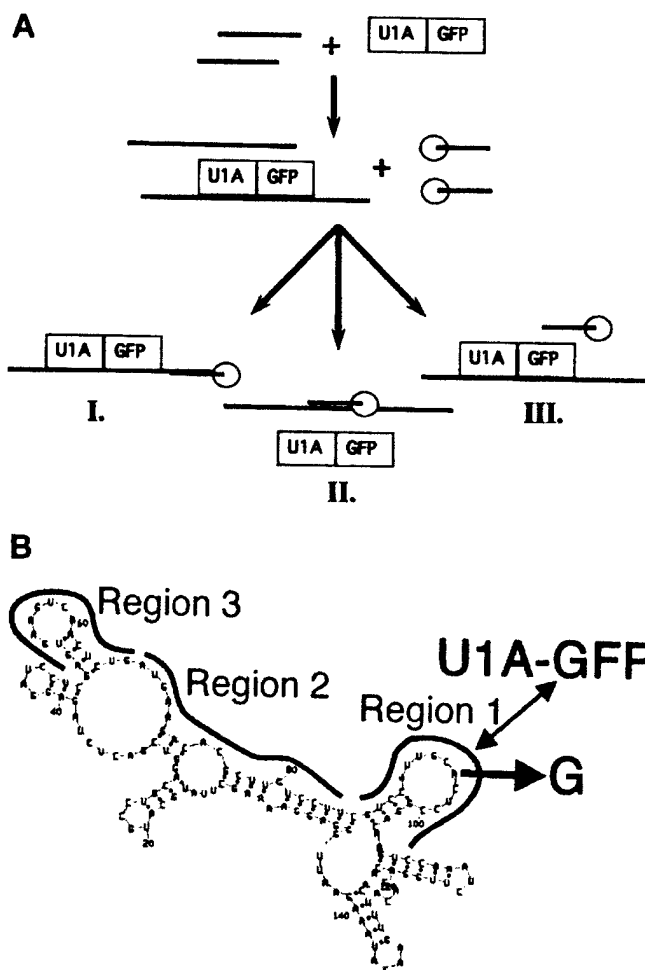


Fig. 1. Experimental design. A, schematic of the experiment. A U1A-GFP-RNA complex is formed and subsequently challenged with oligonucleotide beads. After reaching equilibrium, RNA and protein fluorescence on each microbead are determined by flow cytometry. The experiment can be performed with or without fluorescently labeled RNA. Three scenarios are possible. I, GFP signal is observed indicating the bead is coupled to an oligo complementary to the RNA target but distant from the U1A stem loop recognition site. II, no GFP signal is observed, but the oligo is hybridizing to the RNA. With labeled RNA, the RNA-oligo hybridization is detected. These oligos may be complementary to the U1A binding site. III, beads with neither GFP nor RNA fluorescence are observed, suggesting that these oligos do not hybridize to the RNA. These sequences may be non-complementary to the RNA. B, predicted secondary structure of U1A RNA constructed for these studies as determined by mFOLD (21). U1A-GFP binds to the hairpin derived from U1 snRNA as indicated. An A to G mutation (96A→G) reduces binding 1000-fold. Region 1 is complementary to the binding site. Oligonucleotides complementary to other regions of the RNA, distant from the binding site, are also indicated. These sequences are predicted to hybridize to the RNA while U1A-GFP is binding, allowing the U1A RNA-protein interaction to be observed.

and then incubated with beads to which oligonucleotides complementary to the target RNA have been coupled. In the experiments described here, RNA is labeled with Texas Red, and the protein is a GFP fusion. After reaching equilibrium, the

TABLE I
Oligonucleotides used in this study

Mismatch nucleotides are in lowercase.

Name	Sequence
Oligonucleotides complementary to U1A binding site	
1.20 ^a	TTGTCCGAGTGCAATGGAC
1.17	GTCCGGAGTGCAATGGA
1.15	GTCCGGAGTGCAATG
Oligonucleotides complementary to regions distant from U1A binding site	
2.20	AAGGACAAGCGTGCTTCAT
2.17	GGACAAGCGTGCTTCA
2.15	GGACAAGCGTGCTT
3.17	TCAGAGTTGACTTCACT
3.15	TCAGAGTTGACTTCA
Mismatch oligonucleotides	
2C.20 ^b	AAGGACAACCGTGCTTCAT
2C.17	GGACAACCGTGCTTCA
2C.15	GGACAACCGTGCTT
3C.17	TCAGAGTTGACTTCACT
3C.15	TCAGAGTTGACTTCACT
U1 snRNA oligonucleotides	
4.17	CCCTGCCAGGTAAGTAT
4G.17	CCCTGCGAGGTAAGTAT

^a The first number indicates the sequence that is being targeted, while the second number indicates the oligonucleotide length, e.g. 1.20 is an oligonucleotide complementary to region 1 with 20 nucleotides.

^b The letter indicates a point mutation in the oligonucleotide disrupting hybridization to the RNA.

beads are analyzed in a flow cytometer for protein and RNA fluorescence. Beads are sorted into different categories as illustrated in Fig. 1A. 1) Beads with both GFP and RNA signals represent the RNA-protein interaction. The oligonucleotides on the beads hybridize to the RNA without interfering with protein binding. 2) RNA signal alone indicates RNA hybridization with no protein binding. The oligonucleotides on these beads may be competing with the protein to bind the same RNA sequences. 3) Some beads will have no detectable fluorescent signal. These oligonucleotide-coupled beads contain sequences that cannot hybridize to the RNA including those that are noncomplementary or contain point mutations. To quantitate the data, the mean fluorescence and/or the percentage of beads in the different categories is determined.

Development of the Bead Binding Assay—The system was developed with the human splicing protein U1A binding to the stem loop 2 derived from U1 snRNA. A U1A-GFP fusion protein including the first 94 amino acids of the RNA recognition motif was expressed with an N-terminal His₆ tag and a C-terminal GFP. The histidine tag was proteolytically cleaved to generate functional U1A-GFP. A 145-nucleotide RNA encoding random sequence and including the specific U1A hairpin was designed. The predicted secondary structure is shown in Fig. 1B. Gel shift mobility experiments confirmed that U1A-GFP binds the U1A RNA with a dissociation constant of ~35 nM in 150 mM KCl at 25 °C (data not shown),

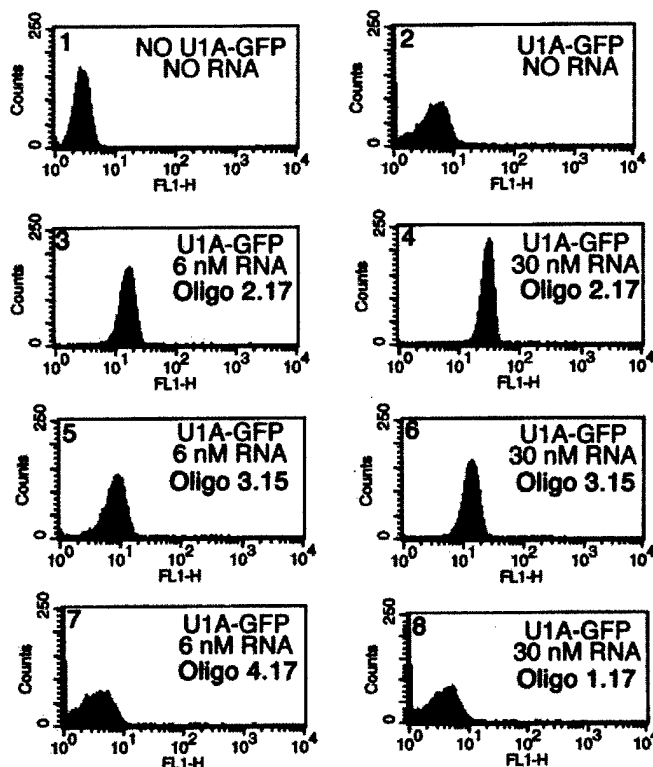


Fig. 2. **Fluorescence is RNA-dependent.** Histograms show the number of beads at different GFP fluorescence intensities. In the presence of RNA and U1A-GFP, the fluorescence intensity of the bead population increases, and a more homogeneous bead population is observed (compare panels 2, 3, and 4). Oligonucleotides targeting the binding site or not complementary to the RNA show nonspecific binding (compare panels 2, 7, and 8).

similar to that reported for the same 94-amino acid fragment in the absence of the GFP (14). Oligonucleotides complementary to different regions of the U1A RNA were designed as illustrated in Fig. 1B and listed in Table I. Oligonucleotides were synthesized with a 5' biotin label and attached to streptavidin beads. Reproducible coupling conditions were devised to ensure similar oligonucleotide concentrations per bead as determined with ³²P-labeled oligos (data not shown). The oligonucleotide concentrations used in the binding assay are estimated to be between 1–2 nM. For all experiments, Dynal 2.8-μm diameter streptavidin beads were used unless otherwise indicated.

After the RNA-protein complex is formed, oligo-coupled beads are added. To reach equilibrium, incubations at 25 °C or 35 °C for longer than 6 h were necessary (data not shown) and typical incubations were at least 14 h. After reaching equilibrium, GFP fluorescence on individual beads was assessed in a flow cytometer. The RNA dependence and specificity of the binding reactions were assessed as follows to ascertain the validity of the approach.

The observed GFP fluorescence on the beads is RNA-dependent as illustrated in Fig. 2A. Background U1A-GFP binding to the beads is low, and the peaks are broad indicative of

a relatively heterogeneous population (Fig. 2, *panel 2*). The observed GFP fluorescence intensity increases with higher U1A RNA concentrations, and the population is more homogeneous as indicated by the narrower peak width (Fig. 2, *panels 3 and 4*). The observed signals were reproducible with different RNA and protein preparations.

The assay accurately distinguishes the U1A RNA binding site. Oligonucleotides complementary to the binding site compete with U1A-GFP for the same RNA sequence and thereby reduce observed GFP fluorescence. Only nonspecific background GFP signal is observed for oligos complementary to the binding site (Fig. 2, compare *panels 2 and 8*). On the other hand, oligos complementary to sequences not part of the U1A binding site show significant GFP signal (oligos 2.17 and 3.15, Fig. 2A, *panels 3–6*). These oligonucleotides are hybridizing to the RNA without interfering with U1A-GFP binding thereby allowing the observation of the RNA-protein interaction. As a control, oligonucleotides not complementary to the RNA show background nonspecific signal (Fig. 2, compare *panels 2 and 7*).

Decreasing the oligonucleotide length lowers the GFP fluorescence intensity. 20-mers, 17-mers, and 15-mers all yield significant fluorescence, while 10-mers complementary to the same region do not (data not shown). Lower fluorescence is consistently observed for 15-mers, such as oligo 3.15, compared with 17-mers, such as oligo 2.17 (Fig. 2, compare *panels 3 and 4 to panels 5 and 6*). These observations are not limited to a particular region of the RNA or sequence.

Discrimination of Oligonucleotide Mismatches—The bead assay discriminates between oligonucleotides that contain mismatches under conditions that preserve RNA-protein interactions. When mismatches in the middle of the complementary sequence are introduced, the oligonucleotide yields significantly lower GFP fluorescence. Mismatch discrimination is not unique to a particular sequence as oligonucleotides complementary to distinct regions show a significant difference in GFP fluorescence (Fig. 3a, compare oligos 2.17 and 2C.17 and oligos 3.17 and 3C.17). Interestingly, unlike 15-mers and 17-mers, 20-mers do not discriminate mismatches as well (Fig. 3a, compare oligos 2.20 and 2C.20 with oligos 2.17 and 2C.17). Also mismatches at the first or second position of either end of the oligonucleotide are not discriminated as well as those in the middle of the sequence (data not shown).

To verify the observed oligonucleotide mismatch discrimination, a compensatory mutation in the RNA was constructed. Binding reactions were prepared with the two different RNAs, and beads coupled to either oligo 2.17 or 2C.17 were added. As observed previously for the wild-type U1A RNA, oligo 2.17 shows significant GFP fluorescence, while oligo 2C.17 does not (Fig. 3b). However, the compensatory mutation in the U1A RNA, 77C→G, creates a mismatch for oligo 2.17 and disrupts the hybridization thereby reducing the observed GFP fluorescence. Meanwhile, significant GFP fluorescence is observed for oligo 2C.17, which is complementary to 77C→G RNA (Fig.

3b). Compensatory RNA mutations and subsequent U1A binding have also been performed with oligos 3.17 and 3C.17 (data not shown). These “rescue” experiments further verify the observed point mutant hybridization discrimination.

The observed mismatch discrimination is enhanced by measuring hybridization through U1A-GFP binding. In the absence of U1A-GFP, poor discrimination between oligonucleotide mismatches is observed (Fig. 3c, compare *panels D and E*). This is consistent with reports of poor hybridization behavior of short oligonucleotide sequence (15, 16). However, in the presence of U1A-GFP, the same oligo 2.17 beads show both a higher GFP and Texas Red fluorescence intensity, while two different mismatches of oligo 2.17 show significantly reduced GFP fluorescence (Fig. 3c). Thus, in physiological conditions, mismatch discrimination is observed by monitoring hybridization through an RNA-protein interaction.

Detection of Specific RNA-Protein Interactions—To determine whether the observed fluorescence is accurately reflecting the U1A RNA-protein interaction, an A to G point mutation in the U1A loop (96A→G) known to disrupt binding was tested (14). This mutation severely reduces the observed GFP signal as shown in Fig. 4A. To quantitate Texas Red and GFP fluorescence, the percentage of the bead population shifted beyond the signals observed for background binding is determined. The quadrants shown in *panels A–C* of Fig. 3c determine the cut-offs to define the bead populations with different combinations of Texas Red and GFP fluorescence. Oligos 2.17 and 3.17 show U1A-GFP signal with wild-type U1A RNA, while only background fluorescence is observed with 96A→G. Importantly both oligonucleotides are hybridizing to the RNA as indicated by significant Texas Red fluorescence suggesting that the lower GFP signal is due to disruption of the protein-RNA complex and not reduced hybridization.

As predicted, oligo 1.17, complementary to the U1A binding site, does not show any significant U1A-GFP binding, similar to the experiments described above. Importantly oligo 1.17 is hybridizing to the RNA at levels similar to other oligonucleotides where GFP fluorescence is observed. This suggests that the low observed GFP fluorescence is due to disruption of the RNA-protein complex and not poor hybridization. These data demonstrate the ability of this system to define sequences important for protein recognition on the RNA by footprinting.

To further demonstrate that the observed GFP fluorescence accurately reflects the RNA-protein interaction, the affinity of the U1A complex was measured (Fig. 4B). A 75 nM dissociation constant in 300 mM KCl at 25 °C is estimated by curve fitting to a Langmuir isotherm, consistent with published data (14). Meanwhile, the 96A→G point mutant shows no significant binding under the same conditions, consistent with its ~1000-fold weaker affinity for this U1A construct (14). Higher nonspecific U1A-GFP binding to the beads causes broader bead population distributions and is probably responsible for the larger error bars observed at higher protein concentrations.

The observed GFP fluorescence is also sensitive to the salt

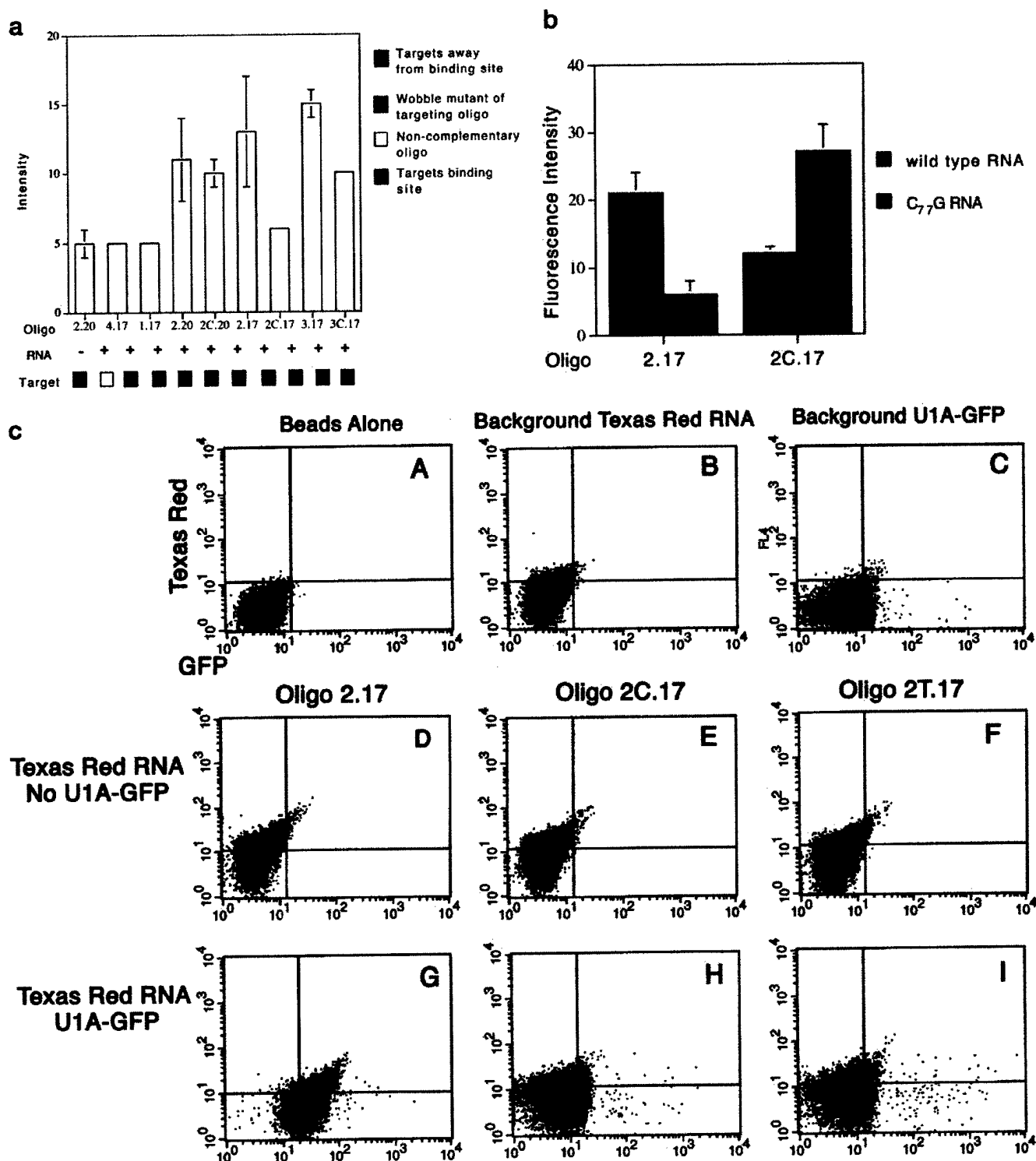
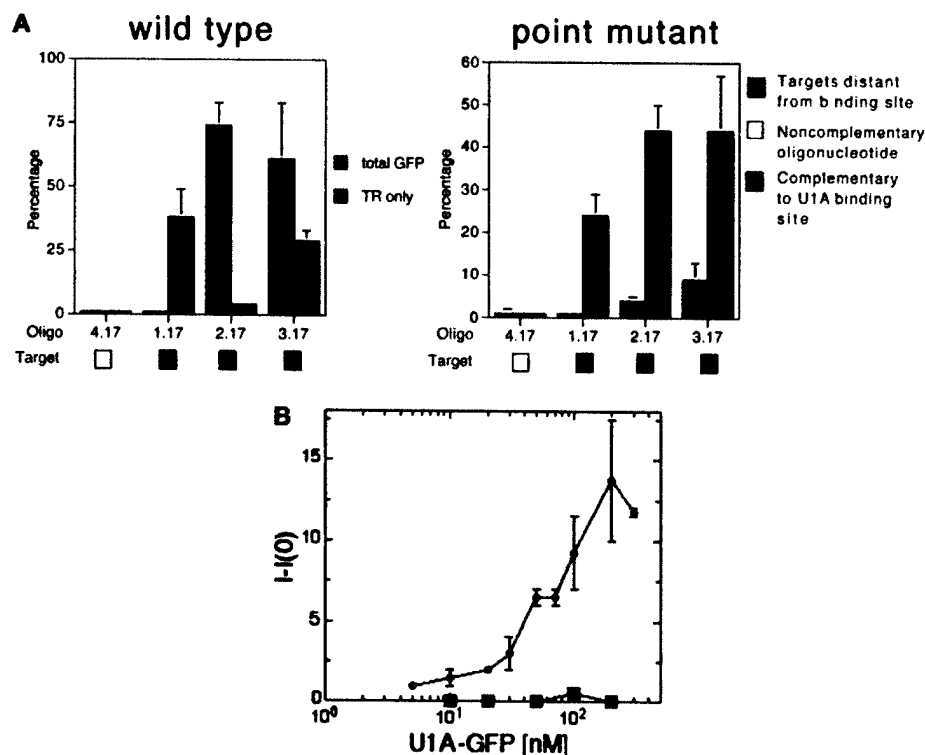


FIG. 3. **Mismatches are discriminated when monitoring hybridization indirectly through U1A-GFP binding.** a, the bar graph shows the mean fluorescence of observed GFP signal of bead populations. Fluorescence intensity is in arbitrary units. Triplicate data are averaged, and the *error bars* represent standard deviations. When the standard deviation is less than 1, no error bar is shown. Binding reactions include 75 nM U1A-GFP and 6 nM U1A test RNA. b, compensatory RNA mutations restore U1A-GFP binding for an oligonucleotide point mutant. Oligo 2.17 hybridizes to the RNA and shows U1A-GFP binding, while 2C.17 does not. 77C→G, which is an exact match to oligo 2C.17, disrupts hybridization to oligo 2.17 and allows hybridization to oligo 2C.17. Binding reactions include 10 nM U1A RNA and 100 nM U1A-GFP. *Error bars* represent standard deviations of duplicates. c, plots show Texas Red fluorescence on the y axis and GFP fluorescence on the x axis. *Panels A–C* show the nonspecific binding of Texas Red RNA and U1A-GFP to the beads. *Panels D–F* illustrate the poor discrimination of 17-mer oligos

FIG. 4. Bead assay accurately reflects the RNA-protein interaction. *A*, percentage of beads shifted above background are tabulated into different categories: beads with Texas Red fluorescence only and beads with GFP fluorescence. U1A-GFP does not bind the 96A→G point mutant RNA, but oligonucleotides still significantly hybridize to the RNA as indicated by the Texas Red fluorescence. *Error bars* represent standard deviations of triplicate experiments. *B*, titration of U1A-GFP into 3 nM RNA generates binding curves. *Squares* represent the 96A→G point mutant, and *circles* represent the wild-type U1A RNA. Background fluorescence (I_0) in the absence of RNA is subtracted for each protein concentration. Binding curves were performed in duplicate and averaged. *Error bars* represent standard deviations. Curve fitting suggests U1A-GFP is binding with a dissociation constant of ~75 nM in 300 mM KCl at 25 °C.



concentration. At higher KCl concentrations, the GFP fluorescence intensity increases by ~25% at each protein concentration. However, the dissociation constant shifts from ~35 to ~75 nM. These observations reflect stronger hybridization and weaker U1A interaction at higher salt concentrations. In sum, these data demonstrate the specificity and affinity of the U1A RNA-protein interaction on beads.

Specific Binding in Mixed Populations—For screening RNA-protein interactions with the bead assay, RNAs will have to be identified from complex mixtures of RNAs. To determine whether total RNA can compete with U1A-GFP binding, yeast RNA was added to the binding reactions. Human U1A does not specifically bind any yeast RNA (17, 18). Only in the presence of the U1A test RNA is GFP fluorescence detectable with oligo 2.17 as shown in Fig. 5A. Oligonucleotides not complementary to the U1A RNA such as oligo 4.17 show fluorescence equivalent to background. This suggests that even in contexts where there might be high nonspecific binding, specific binding is observable. These experiments were performed with larger 5.7- μ m diameter beads because the sensitivity is higher (data not shown). The nonspecific binding on these larger beads is also higher as more GFP fluorescence is observed. However, specific GFP fluorescence is observed at lower concentrations of U1A-GFP compared with the smaller 2.8- μ m microbeads.

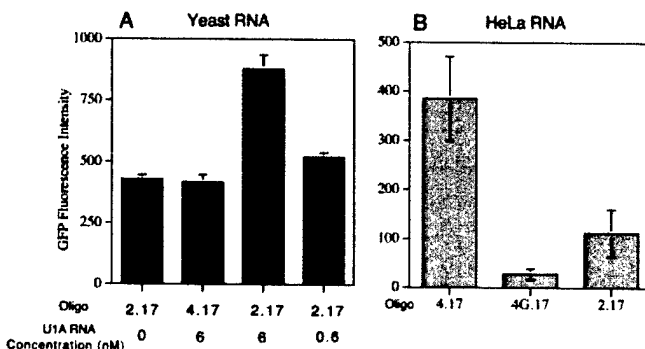


FIG. 5. U1A-GFP binding to the U1A test RNA and to U1 snRNA is detectable in total RNA backgrounds. *A*, oligo 2.17 but not oligo 4.17 detects U1A-GFP signal in 0.1 μ g/ μ l yeast total RNA and 25 nM U1A-GFP. Fluorescence intensities are higher with larger 5.7- μ m beads. Data are in triplicate with standard deviation *error bars*. *B*, conversely oligo 4.17 shows GFP fluorescence in HeLa total RNA. U1A-GFP is binding to snRNA in the total HeLa RNA. Oligo 4.17 also shows higher GFP fluorescence than a mismatch, oligo 4G.17, or a noncomplementary sequence, oligo 2.17. Reactions include 0.16 μ g/ μ l HeLa total RNA (40 μ g total) and 50 nM U1A-GFP. *Error bars* represent standard deviations of triplicates. Background signals were subtracted for this graph.

To determine whether RNAs isolated from RNA preparations can be identified with microbeads, total RNA was isolated from HeLa cells and mixed with U1A-GFP. In HeLa RNA,

binding to the RNA in the absence of U1A-GFP. Oligonucleotide mismatches do not significantly affect the observed Texas Red signal. Only the wild-type oligonucleotide shown in *panel G* gives significant GFP fluorescence with 100 nM U1A-GFP and 30 nM U1A RNA, while the mismatches, in *panels H* and *I*, show some Texas Red signal but no significant GFP fluorescence.

U1A binds to U1 snRNA and its own mRNA. An oligonucleotide complementary to snRNA, oligo 4.17, shows higher GFP fluorescence compared with a mismatch, oligo 4G.17, and a noncomplementary oligonucleotide, oligo 2.17, as illustrated in Fig. 5B. The U1 snRNA concentration is ~1–5 nM in these experiments. The observed GFP fluorescence is HeLa RNA-dependent as the signal varies with HeLa RNA concentration but remains unchanged with increasing equivalent concentrations of yeast RNA. In sum, these data demonstrate that a specific RNA can be identified with microbeads from total RNA.

An additional requirement for screening with oligonucleotide bead libraries is the ability to detect a small percentage of oligonucleotide beads from a large background of beads that do not bind. Initial experiments diluting the oligonucleotide beads 100-fold suggest that at even low oligonucleotide concentrations the RNA-protein complex can be detected. Further experiments diluting oligo 2.17 in a large excess of oligo 4.17 demonstrated that two populations of beads could be differentiated by GFP fluorescence (data not shown). These data also demonstrate that the observed shifts are sufficient to identify an RNA-protein complex.

DISCUSSION

We describe a system to monitor RNA-protein interactions in solution with microbeads using flow cytometry. We demonstrate the versatility of the approach for 1) discriminating between mismatches in the oligonucleotides, 2) mapping protein recognition sites on RNA, 3) differentiating specific and nonspecific binding RNAs, and 4) detecting specific RNAs in complex mixtures. Importantly, specific binding can be detected in high nonspecific RNA backgrounds, and the system can discriminate a nonspecific binding point mutant at a variety of protein and RNA concentrations. Because flow cytometry is used to monitor the fluorescence on distinct oligonucleotide-coupled microbeads, the system is amenable to high throughput, genomic scale screening of RNA-protein interactions.

With the U1A interaction, we have determined the fundamental requirements for using this microbead system for screening RNA-protein interactions. The approximate 2-fold changes in observed GFP fluorescence are sufficient to distinguish specific protein binding to RNA from background. Future versions of the system may have increased sensitivity and dynamic range by using brighter fluorophores and microbeads with higher oligonucleotide densities. Furthermore, because the system is monitoring the protein binding to a distribution of thousands of microbeads, the fluorescence shifts are more significant than just monitoring the bulk signal. Methods to analyze the distributions more quantitatively are being developed. The microbead system also allows for the measurement of relative affinities of a protein for its cognate RNA thereby distinguishing specific and nonspecific binding during screening and thereby reducing false positives. Two

approaches are possible. After initial screening at a particular protein concentration, binding experiments will help distinguish specific and nonspecific binding candidates. Alternatively, screening at different protein concentrations could be performed to determine relatively strong and weak binding interactions.

The microbead-based system described here has a number of advantages over other recently developed RNA-protein screening strategies. It is rapid with few time-consuming manipulations required. Also, unlike many *in vivo* strategies, there are no limitations to the size of the RNA or its basic structural features. It is often difficult to monitor the binding of large RNAs directly because they do not migrate well in gel electrophoresis for mobility shift studies. With the microbead system, the only requirement is that oligonucleotides hybridize to regions distant from the binding site. Most mRNAs have multiple regions accessible to hybridization in physiological conditions (19).

With the system presented here, hybridization is monitored indirectly through the protein fluorescence. To observe protein fluorescence, the oligonucleotides need to hybridize to particular RNAs that are also binding the protein. The combination of observed RNA hybridization and protein fluorescence on a microbead indicates that an RNA-protein complex is present. Thus, many RNAs may be hybridizing to the beads, but only when the fluorescently labeled protein is bound with high affinity to one of these RNAs is positive signal observed. This significantly reduces the nonspecific binding that would be observed in identifying possible RNA targets compared with other strategies that isolate all the RNAs bound to beads. In its current form, the system does not reach equilibrium for hours presumably because of slow hybridization at physiological temperatures. Smaller volumes may help reduce the time required to reach equilibrium.

The mismatch discrimination observed with this bead strategy may allow it to be adopted for single nucleotide polymorphism analysis. Similar to assays such as the invasive cleavage method (20), monitoring hybridization indirectly provides a sensitivity enhancement to observe the subtle effects of mismatches on hybridization. The enhanced mismatch discrimination observed through the protein interaction may be particular to the U1A system. Future studies of other RNA-binding proteins will determine the generality of the observed mismatch discrimination.

For genomic screening, proteins bound to RNA could be challenged with large oligonucleotide coded bead libraries. The coded beads would be sorted in a flow cytometer while monitoring RNA and protein fluorescence to determine which sequences are hybridizing to the RNA while preserving the RNA-protein complex. This information can then be compared with sequenced genomes to determine which RNAs are binding and which sequences may be important for the interaction. Various coding strategies are currently being developed that do not require decoding or very large beads (9–12).

In sum, we have developed a microbead-based system to determine which RNAs may be binding a particular protein as well as which RNA sequences may be important for the RNA-protein interaction. Many applications of the assay are possible including binding in cell extracts, single nucleotide polymorphism analysis, and monitoring the effects of small molecules on RNA-protein complexes. Perhaps the most inviting aspect of this system is to use large coded oligonucleotide bead libraries to probe RNA-protein interactions on genomic scales.

Acknowledgments—We thank Kathy Hall for U1A plasmids, Vicki Losick for constructing 96A→G, and the staff at the Dana-Farber Cancer Institute flow cytometry facility. We thank M. Damelin, C. Stern, P. Zarrinkar, and D. Treiber for critical reading of the manuscript and Jeff Way and Matt Trau for stimulating discussions.

* A portion of this work was supported by grants from the National Institutes of Health and the Department of Defense (to P. A. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by the Claudia Barr Young Investigator program.

§ To whom correspondence should be addressed: Dana-Farber Cancer Inst., SM 922, 1 Jimmy Fund Way, Boston, MA 02115. Tel.: 617-632-5105; Fax: 617-632-5103; E-mail: alex_brodsky@dfci.harvard.edu.

REFERENCES

- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921
- Zhang, B., Kraemer, B., SenGupta, D., Fields, S. & Wickens, M. (2000) Yeast three-hybrid system to detect and analyze RNA-protein interactions. *Methods Enzymol.* **318**, 399–419
- Danner, S. & Belasco, J. G. (2001) T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12954–12959
- Paraskeva, E. & Hentze, M. W. (2000) Translational repression assay procedure: a method to study RNA-protein interactions in yeast. *Methods Enzymol.* **318**, 374–384
- Andrews, L. G. & Keene, J. D. (1999) Identification of specific protein-RNA target sites using libraries of natural sequences. *Methods Mol. Biol.* **118**, 233–244
- Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 14085–14090
- Brown, V., Jin, P., Ceman, S., Darnell, J. C., O'Donnell, W. T., Tenenbaum, S. A., Jin, X., Feng, Y., Wilkinson, K. D., Keene, J. D., Darnell, R. B. & Warren, S. T. (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**, 477–487
- Darnell, J. C., Jensen, K. B., Jin, P., Brown, V., Warren, S. T. & Darnell, R. B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107**, 489–499
- Battersby, B. J., Lawrie, G. A. & Trau, M. (2001) Optical encoding of microbeads for gene screening: alternatives to microarrays. *Drug Discov. Today (HTS Suppl.)* **6**, S19
- Chan, W. C., Maxwell, D. J., Gao, X., Bailey, R. E., Han, M. & Nie, S. (2002) Luminescent quantum dots for multiplexed biological detection and imaging. *Curr. Opin. Biotechnol.* **13**, 40–46
- Gao, Y. C., Jin, R. & Mirkin, C. A. (2002) Nanoparticles with Raman spectroscopic fingerprints for DNA and RNA detection. *Science* **297**, 1536–1540
- Cunin, F. *et al.* (2002) Biomolecular screening with encoded porous-silicon photonic crystals. *Nat. Mater.* **1**, 39–41
- Wilkins Stevens, P., Henry, M. R. & Kelso, D. M. (1999) DNA hybridization on microparticles: determining capture-probe density and equilibrium dissociation constants. *Nucleic Acids Res.* **27**, 1719–1727
- Zeng, Q. & Hall, K. B. (1997) Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability. *RNA* **3**, 303–314
- Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* **22**, 5456–5465
- Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. & Brown, E. L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680
- Tang, J. & Rosbash, M. (1996) Characterization of yeast U1 snRNP A protein: identification of the N-terminal RNA binding domain (RBD) binding site and evidence that the C-terminal RBD functions in splicing. *RNA* **2**, 1058–1070
- Brodsky, A. S. & Silver, P. A. (2000) Pre-mRNA processing factors are required for nuclear export. *RNA* **6**, 1737–1749
- Ho, S. P., Bao, Y., Leshner, T., Malhotra, R., Ma, L. Y., Fluharty, S. J. & Sakai, R. R. (1998) Mapping of RNA accessible sites for antisense experiments with oligonucleotide libraries. *Nat. Biotechnol.* **16**, 59–63
- Wilkins Stevens, P., Hall, J. G., Lyamichev, V., Neri, B. P., Lu, M., Wang, L., Smith, L. M. & Kelso, D. M. (2001) Analysis of single nucleotide polymorphisms with solid phase invasive cleavage reactions. *Nucleic Acids Res.* **29**, E77
- Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940

Analysis of RNA-protein Interactions by Flow Cytometry

Alexander S. Brodsky^{1*}, Angus P.R. Johnston², Matt Trau², and Pamela A. Silver¹

1 Dana-Farber Cancer Institute and
 Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School
 1 Jimmy Fund Way; SM 922
 Boston, MA 02115

2 Centre for Nanotechnology and Biomaterials
 Department of Chemistry
 University of Queensland, St. Lucia 4072
 Brisbane, Australia

* corresponding author, alex_brodsky@dfci.harvard.edu

Abstract

In combination with advances in bead coding technologies, flow cytometry is maturing as a powerful high-throughput approach for analyzing molecular interactions. Applications of this technology include antibody assays and single nucleotide polymorphism mapping. This article describes the recent development of a microbead flow cytometric approach to analyze RNA-protein interactions and reviews emerging bead coding strategies that together will allow genome-wide identification of RNA-protein complexes. The approach is flexible and provides new opportunities for functional genomic studies and small molecule screening.

Key terms

RNA-protein interactions, microbeads, microspheres, flow cytometry, functional genomics

Introduction

The determination of RNA-protein regulatory networks is critical for understanding biological pathways. The role of RNA and RNA-protein interactions in regulating gene expression is becoming more appreciated with each new discovery. RNA-protein interactions are the backbone of many post-transcriptional processes including mRNA stability, splicing, translation, and localization. Determining which RNAs and proteins bind each other remains a challenging goal in the post-genomics era. Many human diseases such as Fragile X [1] and HIV [2] are controlled by proteins interacting with RNAs. Proteins also are in complexes with noncoding RNAs large and small such as 7SK [3,4] and microRNAs [5] to regulate gene expression. Understanding how RNA-protein interactions shape gene expression pathways on genome-wide levels remains unclear.

This review highlights recent advances in technologies to study RNA-protein interactions using genomic and high-throughput methods. In particular, we will focus on the use of microbeads to explore RNA-protein complexes by flow cytometry. These methods could evolve into diagnostic assays and high-throughput screening of pharmacological agents targeted to RNA-protein interactions. We will introduce the assay and discuss aspects of microbeads technology important for the assay such as microbead multiplexing and surface chemistry.

RNA-protein screening approaches

Many assays have been developed to examine nucleic acid protein interactions in vitro including gel mobility shift, footprinting, and filter binding. Hazbun and Fields recently performed a large-scale electrophoretic gel mobility shift assay (EMSA) to monitor DNA binding proteins from pools of GST yeast protein libraries [6]. However, similar to the other biochemical strategies, EMSA requires many manipulations making genome-wide screening labor intensive and difficult. Also, these

approaches require labeling the RNA to monitor binding making it difficult to pick a particular protein and determine the specifically binding RNAs.

A number of genetic methods have been developed for the analysis of RNA-protein interactions. One system that can screen for either RNA binding proteins or for RNA sequences is the three hybrid assay [7]. However, longer RNA sequences cannot be analyzed and certain sequences cause transcription termination [7]. A second genetic strategy is the Translational Repression Assay Procedure (TRAP) in yeast. This strategy works well with hairpin containing RNA binding sites but has yet to be tested with a variety of RNA structures [8]. More recently, phage display methods have been developed with a model system to clone candidate proteins binding to a specific RNA sequence [9]. Genetic methods in mammalian cell lines, such as the Tat-fusion transcriptional activation system [10] and frameshifting assay [11], offer the ability to screen in the presence of potential binding partners. One drawback of these methods is that the complexes are forced to form in particular cellular compartments that may not be the native location. In addition, they often depend on the generation of cDNA libraries that may be biased towards the most abundant messages and would also miss non-coding RNAs such as microRNAs.

Recently, DNA chips have been used to identify RNAs bound to proteins [1,12-14]. This approach is very promising for inspection of RNA-protein interactions on a genome-wide scale. Typically, RNA-protein complexes are immunoprecipitated, the RNA is isolated and analyzed on DNA chips. Alternatively, protein can be prepared on beads and cell extract can be bound to the bead [15]. However, these approaches rely on the ability to preserve stable interactions during immunoprecipitation; many potentially weak interactions may be lost. In addition, RNA binding proteins typically have high nonspecific binding constants leading to the isolation of a mixture of

specific and nonspecific "bound" species, complicating the analysis. Other experiments such as SELEX may be necessary to help determine the specifically binding RNAs [16].

The microbeads assay

A new recently published approach to RNA biochemistry uses flow cytometry and oligonucleotides attached to microbeads [17]. Figure 1 shows a schematic of the assay. A fluorescently labeled protein is bound to fluorescently labeled RNA before being challenged with oligonucleotides attached to microbeads. After hybridization, under conditions that do not disrupt the RNA-protein interaction, the microbeads are sorted and analyzed by flow cytometry. The presence or absence of RNA and protein signals provides information about the binding interaction. RNA-protein interactions can be specifically identified from complex mixtures while simultaneously characterizing some binding properties such as the dissociation constant (K_d). In addition, by probing with a high density oligonucleotide library against RNAs of interest, the binding site could be determined.

The RNA-protein microbead assay was developed with the U1 snRNP model system. U1-GFP was purified and bound to a 150mer RNA. Binding is indicated by GFP fluorescence on the bead population. RNA mutations, oligonucleotide mismatches and K_d 's were measured to demonstrate the specificity of the assay. Single mismatch discrimination of short oligonucleotides was possible when the signal is monitored through the protein binding. Importantly, RNAs could be specifically detected in total RNA isolated from cells. The sensitivity is in the range of other common flow cytometry assays, as picomolar RNA concentrations could be detected. In this format, the assay is accessible to most molecular biology laboratories as it uses common reagents and many facilities have access to flow cytometers.

The microbead assay is an equilibrium binding assay that offers some distinct advantages for the biochemical characterization of RNA-protein complexes. Firstly, protein binding to large RNAs

can be examined. In fact, larger RNAs offer more hybridization targets for the antisense oligonucleotide probes. Using oligonucleotides targeting different regions of the RNA, binding can be monitored across the whole RNA molecule. Also, binding reactions could be performed in the presence of potential cooperative binding partners by using cell lysates or partially purified cell fractions. Because binding can be monitored at different locations across the RNA molecule, similar to footprinting assays, specific and nonspecific sites may be differentiated.

Flow Cytometry

Flow cytometry is a powerful, sensitive and quantitative technology to measure molecular interactions. Flow cytometry has been successfully applied to examine various protein-protein, protein-DNA and DNA-DNA interactions [18,19]. Because it is fluorescence based, it can also be adapted to monitor real-time kinetics and rapid quench studies. Very high sensitivity can be obtained with 10^5 - 10^6 particles/ml and target concentrations in the 10-100 pM range, well below the K_d of most RNA-protein interactions. Because flow cytometry can focus on just the signal on the microbeads and not the unbound molecules in solution, no washing is typically required, saving significant effort. Also, recent advances in coding microbeads (see below) are bringing the power of multiplexing dozens of samples simultaneously to these assays. With the ability to use automatic sample loaders running 2-3 samples per minute, high-throughput plate reading is now feasible. A number of recent reviews highlight the recent technical advances in flow cytometry allowing for high throughput and sensitivity [18-20].

Microbeads vs. microarrays and hybridization

Binding to microbeads instead of microarrays offers a number of potentially significant advantages. Microbeads have proven to be useful for sensitive and rapid bioanalytical assays.

Companies such as Luminex Corp, Lynx Therapeutics and Quantum Dot Corp. have taken advantage of these properties to devise high throughput approaches to immunoassays [21], sequencing [22,23], and SNP mapping [24,25]. Importantly, microbead assays are typically cost-effective, fast, and require minimal sample quantities.

An important advantage of microbeads for the RNA-protein binding assay is the ability to perform binding on a surface that more closely resembles solution conditions. Hybridization on large planar surfaces is limited by mass transport. On the other hand, microbeads offer better diffusion characteristics [26,27]. This leads to significantly improved hybridization kinetics and thermodynamics with microbeads [26,27].

The basic approach of the microbeads assay is also applicable to microarrays. However, nonspecific hybridization at physiological conditions is a difficult requirement of the assay. Because of the demanding hybridization requirements and the relative ease to synthesize oligonucleotides with long linkers to readily available microbeads, the microbead approach offers a simpler assay than with microarrays at this time.

Encoding Strategies

Bead Libraries

Most high throughput bead based libraries use the optical properties of the support as the library code. The exception to this is the approach from Lynx Therapeutics who use non-encoded support beads, and use a series of molecular markers and identifiers [22]. Optical encoding of supports falls into two broad categories. The first (Luminex Corp, Quantum-dot Corp, Illumina, Nanoplex) is based on separately coding each bead and separately synthesizing the target DNA sequence (or other analyte such as RNA, peptide etc.) then attaching each target to a coded bead. The

alternative technique is to directly synthesize the target molecule on a coded bead in a combinatorial fashion and track each synthetic step each bead experiences [28] .

Separate encoding strategy

The methods that use the separate encoding strategy, have a similar strategy to encode the beads. In each case fluorochromes, fluorescent dyes in the case of Luminex [29] and Illumina [30], fluorescent nanocrystals in the case of Quantum-dot corporation [25,31], are incorporated into polystyrene beads by swelling the polystyrene in a solvent and absorbing dyes or nanocrystals into the particles. The bead is then placed in a different solvent, to shrink the bead, trapping the fluorochrome in the bead. The code is formed by varying the concentration and the combination of fluorochromes present in each bead. The code can be read either by a flow cytometer (Luminex Corp, Quantum-dot Corp) or by optic fiber array (Illumina). DNA sequences are synthesized remotely (either separately in an automated DNA synthesizer or in vivo) and attached to the beads using standard ethylenediamine carbodiimide (EDC) coupling chemistry [32] . The separate encoding techniques are useful for small libraries as it is easy to separately synthesize hundreds of different beads and hundreds of target molecules (see table 1). However, there are limitations for larger libraries [33]. To synthesize a library of 100 000 compounds, it requires 100 000 separate coded beads and 100 000 separately synthesized DNA sequences combined in 100 000 coupling reactions (See figure 1A). Automation of this process is possible, however the size of the library is still limited by the number of coded beads that can be formed.

Nanoplex uses metallic rods (instead of spherical particles) with bands of material with different refractive index to form coding system which is very similar to traditional barcodes but on a microscopic scale [34] . The different refractive index is achieved by incorporating different metals into the rods as they are synthesized. Similar to the other separate encoding strategies, library size is

limited by the number of separate reaction vessels required to synthesize the coded support and the analyte. However, unlike the fluorescent coding approach, the barcode can be incorporated over a large number of steps, so the coding system can code for many more sequences than it would be possible to synthesize in the library. At this time, there is no automated, high throughput method of reading these barcodes.

Combinatorial encoding/synthesizing strategy

In the combinatorial method, a set of optically diverse, but distinguishable, particles are used as the support for synthesizing the target DNA (or other target molecules) [35]. The optically diverse set of particles are synthesized using a combinatorial process where beads are split into a number of reaction vessels and varying concentrations of fluorophores such as organic fluorescent dyes or nanoparticles are covalently incorporated into the beads. The beads are then mixed together and the process is repeated for each subsequent dye. This way it is not necessary to synthesize each coded bead individually [28]. Using this method with 6 fluorophores and 8 levels of intensity for each dye, a library of over 250 000 signatures can be made. Still using only 6 fluorophores, but with 16 levels of intensity, a library of over 16 million sequences can be made (see table 1).

Using a flow cytometer and custom designed electronics, beads can be analyzed and sorted according to the particular optical signatures [28]. Each bead has a predetermined sequence that is uploaded to the modified flow cytometer and sort decisions are made according to the sequences that are required for the particular library. The flow cytometer can sort into 4 directions, each direction corresponding to a different nucleoside. After each sort the nucleosides are coupled to the corresponding beads and once coupling is complete, the beads are mixed together and the process is repeated until the oligonucleotide sequences of the required length are synthesized. At the end of the

process, beads with a known optical signature are synthesized with each unique signature corresponding to a different oligonucleotide sequence (see Figure 1B).

Synthesizing libraries in this way requires beads that can withstand the relatively harsh conditions of DNA synthesis. Polystyrene beads are typically not suitable for this process because they swell and leach dye during the synthesis procedure, thus catastrophically altering the optical signature. Therefore specially synthesized silica particles are required [28] .

Surface chemistry

The surface chemistry of the beads plays an important role in the assay. Non-specific binding of proteins to the beads is a larger problem than the non-specific binding of oligonucleotides to the beads, because generally, most surfaces with a large negative charge (such as silica, and polystyrene surfaces) have relatively low non-specific binding of oligonucleotides (because of repulsion of the negatively charged phosphate backbone). However, because proteins have positively and negatively charged regions, it is necessary to have a surface which has little or no surface charge to minimize non-specific electrostatic binding and molecule adsorption[36] . Coating the surfaces with hydrophobic chains (such as alkyl chains) is also not ideal, because many proteins have hydrophobic regions that will also non-specifically bind to the beads [37] . One solution is to add a large excess of a protein (such as cheap and abundant proteins like BSA) that non-specifically bind to the surface of the beads, limiting the non-specific binding of the fluorescently labeled protein. However, there is a limit to the effectiveness of this procedure and it is desirable to have a 'biologically silent' surface that limits the non-specific binding of the proteins.

Much of the surface chemistry developed for protein-chips can be applied to bead surfaces. Polyethylene glycol and oligo-ethylene glycol surfaces have been used to minimize the non-specific binding of proteins [37] to silica substrates.

Surface density of the probes also plays an important role in the assay. Clearly the higher the number of probes on the beads results in a higher signal, however overloading the surface introduces problems. It is possible to load in excess of 100 million target sequences on a single bead, but at this very high surface density, steric hindrance can effect the hybridization of target DNA to the beads. Also, it can lead to false hybridization events, where one target DNA strand hybridizes to multiple probe strands on the bead [38]. Similar findings have been observed on arrays [39].

Towards High-throughput screening and genomics

This review describes the recent development of a versatile flow cytometry approach to examine RNA-protein interactions. The emerging bead-coding and surface chemistry technologies described herein in combination with novel assays such as the microbead RNA-protein assay will lead to new small molecule and genomic screens. Because of the versatility and flexibility of flow cytometry and the RNA-protein assay described herein, many variations are possible including defining the binding spectrum of a particular RNA binding protein, screening a protein library for binding to a specific RNA, or discovering small molecules that inhibit a RNA-protein interaction. With the increasing understanding of the importance of RNA-protein interactions in human disease and development, we expect these promising technologies to contribute significantly.

References

1. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, Jin X, Feng Y, Wilkinson KD, Keene JD, et al.: **Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome.** *Cell* (2001) **107**:477-487.
2. Frankel AD, Young JA: **HIV-1: fifteen proteins and an RNA.** *Annu Rev Biochem* (1998) **67**:1-25.
3. Nguyen VT, Kiss T, Michels AA, Bensaude O: **7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes.** *Nature* (2001) **414**:322-325.
4. Yang Z, Zhu Q, Luo K, Zhou Q: **The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription.** *Nature* (2001) **414**:317-322.

5. Dostie J, Mourelatos Z, Yang M, Sharma A, Dreyfuss G: **Numerous microRNPs in neuronal cells containing novel microRNAs.** *Rna* (2003) **9**:180-186.
6. Hazbun TR, Fields S: **A genome-wide screen for site-specific DNA-binding proteins.** *Mol Cell Proteomics* (2002) **1**:538-543.
7. Zhang B, Kraemer B, SenGupta D, Fields S, Wickens M: **Yeast Three-Hybrid System to Detect and Analyze RNA-Protein Interactions.** *Methods in Enzymology* (2000) **318**:399-419.
8. Paraskeva E, Hentze MW: **Translational repression assay procedure: a method to study RNA-protein interactions in yeast.** *Methods Enzymol* (2000) **318**:374-384.
9. Danner S, Belasco JG: **T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries.** *Proc Natl Acad Sci U S A* (2001) **98**:12954-12959.
10. Landt SG, Tan R, Frankel AD: **Screening RNA-binding libraries using Tat-fusion system in mammalian cells.** *Methods Enzymol* (2000) **318**:350-363.
11. Kollmus H, Hauser H: **Frameshifting assay to characterize RNA-protein interactions in eukaryotic cells.** *Methods Enzymol* (2000) **318**:363-374.
12. Andrews LG, Keene JD: **Identification of specific protein-RNA target sites using libraries of natural sequences.** *Methods Mol Biol* (1999) **118**:233-244.
13. Tenenbaum SA, Carson CC, Lager PJ, Keene JD: **Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays.** *Proc Natl Acad Sci U S A* (2000) **97**:14085-14090.
14. Hieronymus H, Silver PA: **Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery.** *Nat Genet* (2003) **33**:155-161.
15. Trifillis P, Day N, Kiledjian M: **Finding the right RNA: identification of cellular mRNA substrates for RNA-binding proteins.** *Rna* (1999) **5**:1071-1082.
16. Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB: **Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function.** *Cell* (2001) **107**:489-499.
17. Brodsky AS, Silver PA: **A Microbead-based System for Identifying and Characterizing RNA-Protein Interactions by Flow Cytometry.** *Mol Cell Proteomics* (2002) **1**:922-929.
18. Iannone MA: **Microsphere-based molecular cytometry.** *Clin Lab Med* (2001) **21**:731-742.
19. Sklar LA, Edwards BS, Graves SW, Nolan JP, Prossnitz ER: **Flow cytometric analysis of ligand-receptor interactions and molecular assemblies.** *Annu Rev Biophys Biomol Struct* (2002) **31**:97-119.
20. Nolan JP, Sklar LA: **The emergence of flow cytometry for sensitive, real-time measurements of molecular interactions.** *Nat Biotechnol* (1998) **16**:633-638.
21. Vignali DA: **Multiplexed particle-based flow cytometric assays.** *J Immunol Methods* (2000) **243**:243-255.
22. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, et al.: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays.** *Nat Biotechnol* (2000) **18**:630-634.
23. Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, Mao JJ, Luo S, Kirchner JJ, Eletr S, et al.: **In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs.** *Proc Natl Acad Sci U S A* (2000) **97**:1665-1670.
24. Taylor JD, Briley D, Nguyen Q, Long K, Iannone MA, Li MS, Ye F, Afshari A, Lai E, Wagner M, et al.: **Flow cytometric platform for high-throughput single nucleotide polymorphism analysis.** *Biotechniques* (2001) **30**:661-666, 668-669.

25. Xu H, Sha MY, Wong EY, Uphoff J, Xu Y, Treadway JA, Truong A, O'Brien E, Asquith S, Stubbins M, et al.: **Multiplexed SNP genotyping using the Qbead system: a quantum dot-encoded microsphere-based assay.** *Nucleic Acids Res* (2003) **31**:e43.
26. Henry MR, Wilkins Stevens P, Sun J, Kelso DM: **Real-time measurements of DNA hybridization on microparticles with fluorescence resonance energy transfer.** *Anal Biochem* (1999) **276**:204-214.
27. Wilkins Stevens P, Henry MR, Kelso DM: **DNA hybridization on microparticles: determining capture-probe density and equilibrium dissociation constants.** *Nucleic Acids Res* (1999) **27**:1719-1727.
28. Battersby BJ, Lawrie GA, Johnston AP, Trau M: **Optical barcoding of colloidal suspensions: applications in genomics, proteomics and drug discovery.** *Chem Commun (Camb)* (2002):1435-1441.
29. Fulton RJ, McDade RL, Smith PL, Kienker LJ, Kettman JR, Jr.: **Advanced multiplexed analysis with the FlowMetrix system.** *Clin Chem* (1997) **43**:1749-1756.
30. Walt DR: **Techview: molecular biology. Bead-based fiber-optic arrays.** *Science* (2000) **287**:451-452.
31. Han M, Gao X, Su JZ, Nie S: **Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules.** *Nat Biotechnol* (2001) **19**:631-635.
32. Walsh MK, Wang X, Weimer BC: **Optimizing the immobilization of single-stranded DNA onto glass beads.** *J Biochem Biophys Methods* (2001) **47**:221-231.
33. Battersby BJ, Grondal L, Lawrie GA, Trau M: **Colloids for encoding chemical libraries: Applications in biological screening.** In *Colloids and Colloidal Assemblies*. Edited by Caruso F: Wiley; 2002.
34. Nicewarner-Pena SR, Freeman RG, Reiss BD, He L, Pena DJ, Walton ID, Cromer R, Keating CD, Natan MJ: **Submicrometer metallic barcodes.** *Science* (2001) **294**:137-141.
35. Trau M, Battersby BJ: **Novel Colloidal Materials for High-Throughput Screening Applications in Drug Discovery and Genomics.** *Advanced Materials* (2001) **13**:975-979.
36. Song J, Choi H, Moeser G, Laibinis P: **Preparation of protein- and cell-resistant surfaces on SiO₂ surfaces by polyther-grafted poly(acrylic acid) thin films.** *Polymer Preprints* (2003) **44**:189-190.
37. Lee SW, Laibinis PE: **Protein-resistant coatings for glass and metal oxide surfaces derived from oligo(ethylene glycol)-terminated alkytrichlorosilanes.** *Biomaterials* (1998) **19**:1669-1675.
38. Guo Z, Gatterman MS, Hood L, Hansen JA, Petersdorf EW: **Oligonucleotide arrays for high-throughput SNPs detection in the MHC class I genes: HLA-B as a model system.** *Genome Res* (2002) **12**:447-457.
39. Peterson AW, Heaton RJ, Georgiadis RM: **The effect of surface probe density on DNA hybridization.** *Nucleic Acids Res* (2001) **29**:5163-5168.
40. Hacia JG: **Resequencing and mutational analysis using oligonucleotide microarrays.** *Nat Genet* (1999) **21**:42-47.

Reference comments

of special interest

of outstanding interest

11. The first use of microarrays to analyze coimmunoprecipitating mRNAs.
13. Perhaps the most complete microarray analysis of coimmunoprecipitating bound mRNAs published to date.
16. The description of the novel microbeads assay forming the basis of this review.
24. The authors describe the use and advantages of a quantum dot based optical code.
25. Experiments comparing microbeads' hybridization kinetics to solution.
26. Experiments comparing microbeads' hybridization thermodynamics to solution.
27. The description of the combinatorial coding technique
36. The authors describe the advantages of using ethylene glycol surfaces to minimize non-specific binding.

Figure Legends

Figure 1. Schematic of the experiment. A fluorescently labeled RNA Binding Protein (RBP)-RNA complex is formed and subsequently challenged with oligonucleotide beads. After reaching equilibrium, RNA and protein fluorescence on each microbead is determined by flow cytometry. The experiment can be performed with or without fluorescently labeled RNA. Three scenarios are possible: I.) RNA and protein fluorescence signal is observed indicating the bead is coupled to an oligo complementary to a RNA molecule that is binding the RBP. II.) No protein fluorescence signal is observed but the oligo is hybridizing to the RNA. With labeled RNA, the RNA-oligo hybridization is detected. These oligos may be complementary to the RBP binding site and compete for RBP binding. III.) Beads with neither protein nor RNA fluorescence suggesting that these oligos do not hybridize to the RNA. These sequences may be noncomplementary to the RNA.

Figure 2. Comparison of encoding techniques. A: Separate encoding strategy where beads are individually coded in separate reaction vessels and the oligonucleotides are individually synthesized remotely. The oligonucleotide is coupled to the bead using standard EDC chemistry. B: Combinatorial encoding strategy where silica particles are coded using a split and mix process using varying concentrations of dyes. Using a customized flow cytometer the particles are sorted into four reaction vessels (one for each base) according the predetermined parameters. The process is repeated until oligonucleotides of the required length are synthesized.

5/4/03

Acknowledgements

ASB is supported by the Claudia Adams Barr Young Investigator program and a NHGRI Genome Scholar award.